Acta Universitatis Wratislaviensis No 4131
Anglica Wratislaviensia LX, Wrocław 2022
https://doi.org/10.19195/0301-7966.60.8

Margaret Niznikiewicz
ORCID: 0000-0003-3191-2005
Harvard Medical School
Boston VA Research Institute

How Communication Happens—Where Physical Properties and Meaning Meet in the Brain: Evidence from Semantic, Prosodic and Face Processing Studies

Abstract: Communication in social groups, especially in human societies, is predicated on efficient decoding of physical properties of auditory and visual signals into messages. In this brief overview, I will discuss processes that lead to our experience of receiving a message focusing on semantic, prosodic and face processing operations. In spite of the fact that we experience such messaging as nearly instantaneous, it involves complex interactions between multiple brain regions that support processes involved in communication. In the course of such interactions neural operations analyze a physical signal, extract its features into abstract representations and assign meaning to them. Furthermore, abnormalities in these processes, brought about by either structural or functional deficits, result in profound cognitive difficulties that often manifest as clinical symptomatology. This chapter discusses in some detail which brain networks make social communication possible, as well as the consequences of their abnormalities.

Keywords: communication, language, physical properties, cognition, brain regions, brain networks

Communication is one of the basic tools of building social groups. While typically it is understood as an exchange of ideas and information, it is much more than that. Across all species, communication is more than language. It is a system of signs that carry meaning which not only includes information about the world, but also reflects attitudes and feelings towards that world, the speaker—or the self—and the receiver—or the other. Thus, communication can be defined as an exchange of semantic, emotional, and socially-relevant information using multiple channels of communication and biologically-based as well as socially- and culturally-agreed-upon semiotic signs. Furthermore, what is often experienced as an effortless exchange of ideas about, and information on, a given topic, is in fact a complex process where physical properties of a given signal are analyzed by specialized brain regions, assigned meaning, integrated with information from other sensory modalities and brain systems, and interpreted by the brain to result in a message. In this chapter, I will discuss how the physical properties of human speech and body language, especially facial features, result in a rich repertoire of signs which allow for effective communication, and which brain regions are involved in translating these physical features into a message. I will also discuss what happens when the complex dance between the different brain systems breaks down.

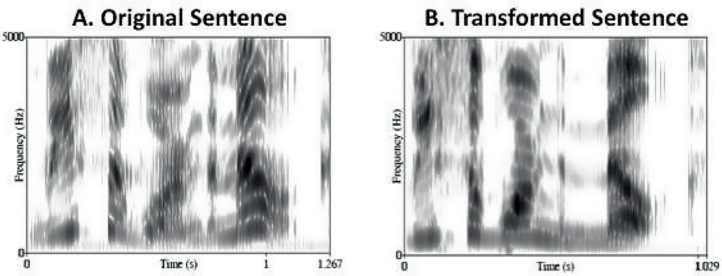

Figure 1: An example of a broad band spectrogram for a sentence “Lisa warmed the milk” (happy prosody): A. with semantic information preserved and B. with semantic information removed. Note that while the two spectrograms differ since their constituent sounds were manipulated, their pitch contours remain the same (spectrograms for single words look similarly)

Until relatively recently, the way the physical properties of a signal carrying a message are analyzed by the brain, so that we perceive them as a meaningful message, was poorly understood. However, advances in several imaging methodologies have changed that. Event-related potential (ERP) and functional magnetic imaging (fMRI) methodologies have proven especially fruitful in unlocking the mysteries of how the brain processes physical properties of a signal to construct meaning.

I will use evidence from both ERP and fMRI studies to discuss how information from a speech signal, and from facial expressions, is interpreted by the brain. ERP and fMRI studies can be thought of as complementary methodologies: while ERPs have an excellent, millisecond-range temporal resolution which allows tracking neural events as they happen in real time, fMRI allows for identifying brain regions involved in these neural events.

In a typical ERP experiment, various stimuli, such as sounds, single words, sentences, or pictures, are presented to participants, either auditorily or visually, while an EEG is recorded. It is then possible to time-lock the EEG signal to the onset of a given stimulus, and thus analyze how that signal changes in response to stimulus presentation. Given the fact that these EEG signals reflect electrical responses generated by brain regions involved in a given type of information processing, we can effectively analyze, and visualize, how the brain processes information with millisecond accuracy. The resulting event related potentials (ERPs) that reflect neural processes associated with changes in electrical voltage over time, are characterized by components which have been demonstrated to index specific cognitive operations. The adopted naming convention for these components, such as P100 or N100, reflects whether the component has a positive- (P) or negative-going (N) valance, measured as amplitude within microvolt range, while the numbers following the P/N, such as N100 or P200 reflect a latency, measured in milliseconds, at which these components are recorded after a participant is exposed to a given stimulus. Early ERP components, such as N100 and P200, reflect mostly sensory processes, while later components, e.g., P300 and N400, reflect primarily more complex processes such as working memory (P300) and language operations (N400). As will be discussed below, most cognitive processes are completed within 1 min, with most processing completed within 500 msec.

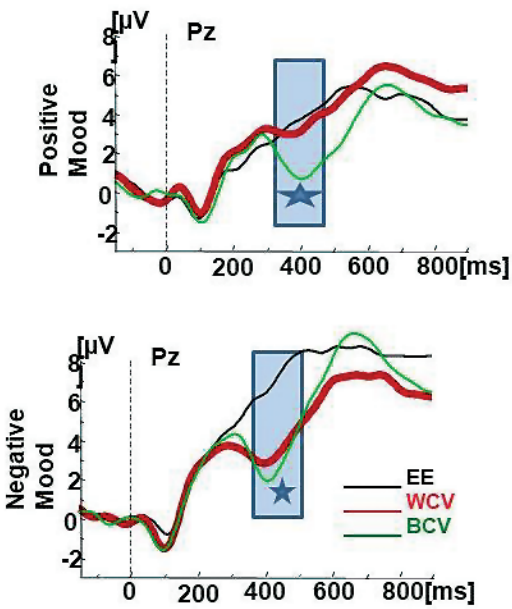

Figure 2: Grand average waveforms showing N400 component to expected endings (EE), within category violations (WCV), and between category violations (BCV)

FMRI allows for detecting differences in blood-oxygen-level-dependent (BOLD) signals. The technique relies on the fact that oxyhemoglobin and deoxyhemoglobin have different magnetic properties which allows for distinguishing between neural tissues preferentially saturated with oxygen in response to cognitive task demands, and those brain regions which are not active in that task. Ultimately, it is possible to create activation maps of the brain that help identify brain regions implicated in a given cognitive operation.

1. Speech signal

All human speech reaches our ears as waves of sounds. They are complex auditory signals with several properties, including pitch, loudness, timbre and carrying a wealth of information coded in its acoustic energy structure organized in acoustic bands called formants (see Figure 1); these acoustic features carry not only semantic information, but also prosodic information which, for speakers of a given language and users of a culture, organizes a sound stream into packets of information and informs about attitudes and emotions.

2. Speech signal, semantics, and the brain

One of the most important functions of a speech signal is to convey a semantic message. As illustrated in Figure 1, a complex acoustic signal brings meaningful information to users of a language, from simple statements such as "Lisa warmed the milk" to more complex messages like "They made a lot of money" and "International peace relies on trust". While fluent speakers of a language experience these informational exchanges as instantaneous, the processes that lead to our holistic experience of receiving a message are complex. It has been demonstrated that the structure and acoustic properties of a speech signal are analyzed by a distributed network of brain regions that include temporal, frontal and parietal cortices (Friederici; Friederici and Gierhan; Chai et al.). As mentioned above, the processes that lead to our 'instantaneous' comprehension, from exposure to a speech signal to our knowing what was said, take under one second, with most of the message, if not all of it, decoded within 400–600 msec. These processes happen both sequentially and in parallel in the sense that regions involved in sensory processing are impacted via feedback loops by abstract lexical, linguistic and semantic information. The superior temporal gyrus has been identified as an interface between brain regions involved in the analysis of physical properties of a speech signal and those involved in the processing of meaning (Bhaya-Grossman and Chang). The STG includes populations of neurons that encode both for the acoustic-phonetic and for phonological properties resulting in abstract representations. It is believed that anterior STG is more involved in acoustic-phonetic analyses, while middle and posterior STG are involved in categorization processes which allow assigning classes of sounds to a given phonological category. Furthermore, these early processes are influenced by feedback from brain regions encoding for lexical and semantic contextual information including from supramarginal gyrus and middle and superior frontal gyri (MFG and SFG) (e.g., Getz and Toscano). Parietal regions, especially the angular gyrus and inferior parietal lobe, are involved in meaning construction (e.g., Hagoort).

It is important to note that these language operations interact with attention, working and long-term memory, allowing for both efficient and flexible extraction of information from an acoustic signal and associating it with context-predicated meaning (e.g., Hagoort). Not only does a speech signal need to be decoded into a semantic message, but this message will be further shaped by other cognitive, non-linguistic factors, as executed by neural networks supporting such operations, including theory of mind (Mar; Mars et al.), and mirror neuron systems (e.g., Rizzolatti and Sinigaglia), and further impacted by emotional states. Theory of mind and mirror neuron systems impact semantic meaning by allowing for inferences about a speaker's intentions, generating hypotheses about what he/she really wanted to say, thus impacting our ultimate understanding of a message.

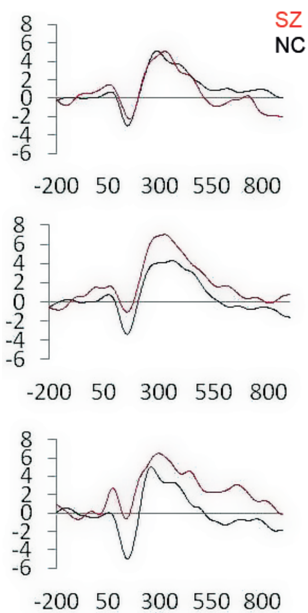

Figure 3: Grand averages in 16 schizophrenia (SZ) and 15 healthy control (NC) subjects in the single words prosody with semantics condition

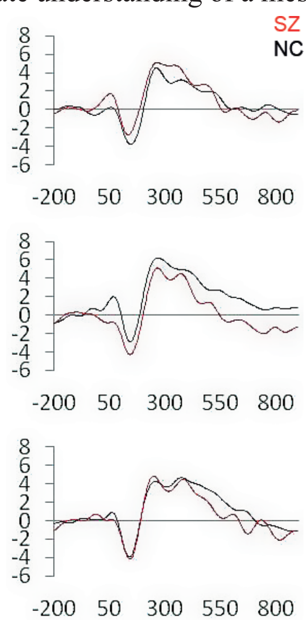

Figure 4: Grand averages in 16 schizophrenia (SZ) and 15 healthy control (NC) subjects in the single words, pure prosody condition

While theory of mind and mirror neuron systems allow for generating predictions about a speaker's intentions, mood and emotional states can further modulate how we receive and understand what is being said. For example, in a study that examined the effects of context and mood on comprehension (Pinheiro et al., "Interactions"; Pinheiro et al., "Abnormal"), we used a two-sentence paradigm where each sentence separately always made sense, but when read together, the second sentence either violated the context by including an incorrect but somewhat plausible word ending or an entirely implausible word ending given a previous context (e.g., *The Joneses made a lot of money and decided to move to a wealthy suburb. There, they bought a large mansion* (expected ending (EE)/*a large apartment* (within category violation)/*a tep*ee (between category violation). We used an ERP methodology and an N400 ERP potential, one of the best documented indices of semantic processes (Kutas and Federmeier, "Electrophysiology" and "Finding"), to examine the effects of induced mood on the processing of within and between category violations. Using carefully-calibrated International Affective Picture System (IAPS) (Lang, Bradley and Cuthbert) stimuli, we induced either a positive or a negative mood in the study participants. The ERP results suggested that when participants were in a positive mood, they treated mild (within category) violations as if they were the correct ending (see Figure 2, top panel). However, when they were in a negative mood, they treated these same mild violations as if they were strong (between category) violations, as indexed by the N400 ERP (see Figure 2, bottom panel). This result is in keeping with these observations, which suggest that people are more creative and make more errors when they are in a positive mood, and they are less creative and make fewer errors when they are in a negative mood.

Together, these results illustrate how a string of noises made by users of a language system carries our thoughts and opinions about the world around us and other people in that world. Deep philosophical, scientific insights, and gossip are all possible due to a network of brain regions which decode this auditory signal into electrical/chemical signals and assign it a meaning in concert with a set of brain regions which situate this meaning within a social context. However, as some early attempts into synthetic speech generation illustrate, no speech is really complete, or indeed possible to understand, without prosody.

3. Speech signal, prosody and the brain

All speech is spoken with prosody, either neutral or emotional. Acoustic features that contribute to prosody are fundamental frequency (F0), loudness, and voice timbre, with F0 believed to be most informative in terms of a tone of voice perceived. As Figure 1 demonstrates, speech signals carrying both semantic and prosodic, and prosodic only information, have their unique acoustic signatures.

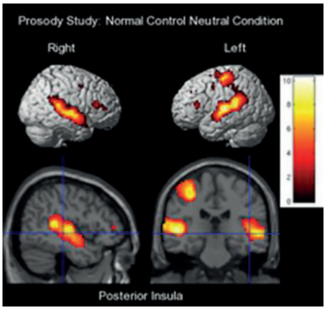

Figure 5: Neutral prosody in healthy volunteers, in sentences which include both semantic and prosodic information: FEW: .001 corrected

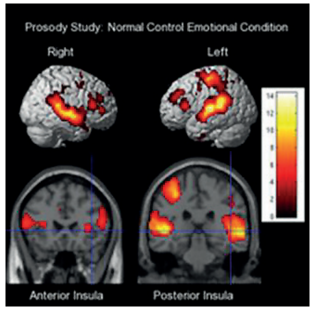

Figure 6. emotional prosody in healthy volunteers, in sentences including both semantic and prosodic information: FEW: .001 corrected

Language users, in the course of neurodevelopment, learn to assign emotional meaning to specific frequency ranges (e.g., Liu et al., "How Tone"). For example, Leitman et al. identified 378 Hz carrier frequency and 169 Hz modulation depth as associated with happy prosody by the study participants, while 178 Hz carrier frequency and 23 Hz modulation depth was recognized as a sad frequency. In English, shifts in F0 occur in the first 300 msec from the onset of an utterance (Kotz and Paulmann; Paulmann and Kotz).

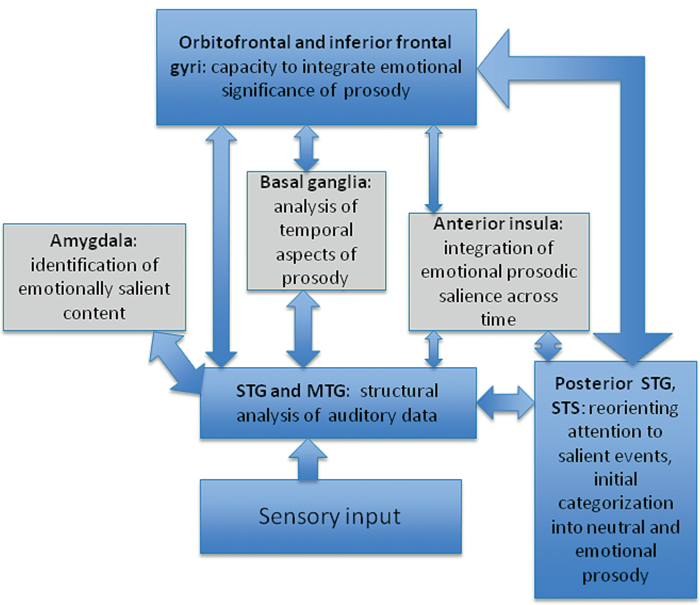

Figure 7: A model of prosody processing according to recent literature findings. Note that the model presented includes both the regions essential for prosody processing (in blue) and regions that have modulatory influence on the essential regions (in gray)

Using ERP methodology, it is possible to examine brain responses to prosodic information, both in utterances with semantic content, i.e., in 'normal' speech, and in those where semantic information was edited out in order to have the listener focus on prosodic information only (see Figure 1). Given the ERP methodology's excellent temporal resolution, it is possible to know exactly when, and how, an acoustic signal is processed by the brain. The results of such ERP experiments demonstrate that the differential neural response to neutral and emotional prosody takes place within the first 200 msec after the onset of the auditory stimulus (Pinheiro et al., "Interactions"; Pinheiro et al., "Abnormal"; Liu et al., "Electrophysiological"). They also allow for several additional observations: they suggest the sensory features of an utterance are rapidly processed within 100 msec after a word/sentence onset, as indexed by N100 component (see Figures 3 and 4), often referred to as the first stage of processing and indexing extracting sensory information from an acoustic signal. It is followed by a P200 component that is associated with classification processes facilitating a rough stimulus appraisal (Garcia-Larrea, Lukaszewicz and Mauguiere; Crowley and Colrain), and its amplitude increases with increased cognitive effort (Lenz et al.). Both N100 and P200, and the processes they index, are impacted by attentional and working memory operations such that the availability of these resources can impact the fidelity of initial operations encoding an adutory signal into more abstract, higher-order representations.

Several studies, using both ERP and fMRI methodology, identified a network of brain regions involved in emotion recognition from voice. These studies also related ERP findings to fMRI results by establishing which brain regions were associated with specific ERP components.

In fMRI studies examining prosody processing in sentences with semantic information, a network of regions active during prosody processing included the temporal, parietal and frontal regions. Notably, more extensive activation was present in studies examining emotional than neutral prosody (see Figures 5 and 6).

Together, these studies allowed for constructing a theoretical framework which describes (see Figure 7) how prosody, i.e., information about our emotions and attitudes is analyzed by the brain. The proposed model consists of a network of both cortical and subcortical areas organized hierarchically (Paulmann and Kotz; Schirmer and Kotz; Leitman et al.). The initial acoustic processing (first stage) takes place within the superior temporal gyrus (STG) and is followed by the categorization of the auditory signal into emotional and non-emotional (second stage) which takes place in posterior region of superior temporal gyrus (pSTG), middle temporal gyrus (MTG) (Okada et al.) and superior temporal sulcus (STS). Recent evidence suggests that sensorimotor cortices are also involved in emotion recognition regardless of sensory modality. The amygdala is involved in initial detection of emotional valence of prosody. The output of these initial operations is further processed by orbitofrontal (OFG) and inferior frontal gyrus (IFG), where emotional meaning is associated with the acoustic signal. It has been also demonstrated that the anterior insula and the superior (SFG) and middle frontal gyrus (MFG) are involved in the processing of emotional prosody and, in cases where the processing is more difficult because of the ambiguity of the signal or contradictory evidence from the vocal and visual channels, both dorsal and ventral parts of anterior cingulate are also involved (Kanske, Plitschka and Kotz). It is of note that ERP and fMRI methodologies not only provide complementary temporal and spatial information about cognitive processes under investigation, but also can provide information in terms of which processing stages can be reflected in each methodology. For example, Paulmann, Seifert and Kotz conducted an ERP study where they tested individuals with orbitofrontal lesions associated with difficulties in recognizing prosody: N100 and P200 in these patients were normal, but they were not able to distinguish between different prosody types because of the orbitofrontal dysfunction.

Finally, similarly to how semantic information is further impacted by non-linguistic inputs, in addition to the structures described above, humans understand emotions due to two complementary systems: a ‘mirror’ system involved in matching motor acts to corresponding motor representations in an observer (Niedenthal et al.) and mentalizing system involved in representation of the mental states of others (theory of mind) (Mitchell and Phillips), which is active when study participants are asked to make explicit judgments about emotional states of others (Spunt and Adolphs, “A New Look” and “The Neuroscience”).

4. When information is presented from faces and voices at the same time

In a typical conversation, spoken words are exchanged between two or more speakers and further modulated by extra-linguistic factors to arrive at the meaning of what is being said. Both semantic information and prosody contribute to an understanding of both a message and an intention: for example, the same simple message of ‘He arrived’ may be delivered with joy, fear or disgust and thus ‘mean’ different things. However, with the exception of phone conversations, most human speech exchanges happen face to face, either in real life or over electronic mediums, and these face-voice interactions add further complexity to an ultimate meaning of a message.

Again, the messages from these two channels: the auditory and the visual one, arrive to us as sounds and images that we learn to associate with specific meanings in the course of culturally mediated neurodevelopment. I have discussed briefly how an auditory stream is processed by the brain to deliver a message. A visual signal that carries information about the human face is equally critical for a full understanding of what a given communication is about.

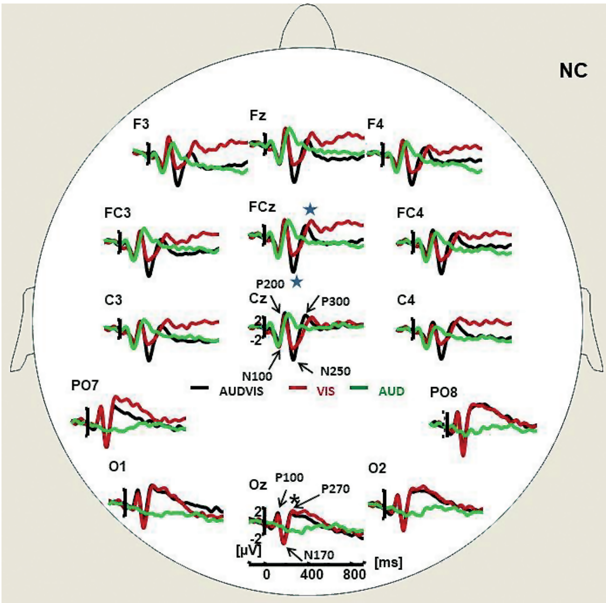

Figure 8: ERP responses in healthy volunteers (NC) in a neutral emotion condition, to faces only (VIS), voices only (AUD) and face and voice stimuli presented simultaneously (AUDVIS)

5. How face information is processed by the brain

As in the case of speech processing, faces are recognized rapidly by dedicated brain regions. A number of studies demonstrated that faces are processed holistically within the first 200 msec after an exposure to a face, with the N170 ERP component believed to uniquely respond to faces (e.g., Yovel; Liu et al., "Emotional"; Allison et al.) and evidence of sensitivity to faces within 100 msec of seeing a face. fMRI and simultaneous fMRI-ERP studies suggest that brain regions involved in face processing include occipital face area (OFA), which has been associated with the 100 msec ERP response, and fusiform face area (FFA),and posterior superior temporal sulcus (p-STS) associated with the N170 ERP response (Iidaka et al.; Sadeh and Yovel; Yovel). These brain regions seem to 'specialize' in face processing relative to other objects, and it has been demonstrated that they are most active when a face is presented. Furthermore, there is evidence that the preferential processing of faces in these brain regions develops in early infancy as a result of social interactions and exposure to faces (Powell, Kosakowski and Saxe).

6. How face and voice are processed together

The simultaneous processing of information from both voice and face is, in fact, a subject of intense studies and debate and it is not entirely understood how such neurocognitive processes happen (Perrodin et al.). It has been suggested that superior temporal sulcus (STS) is sensitive to inputs from both voice and face and is a brain region that supports the integration of visual and auditory signals (Chandrasekaran and Ghazanfar). In addition, fMRI study by Kreifelts et al. suggested a role of the bilateral posterior STG and right thalamus in the integration of auditory and visual signals. These findings are supported at least partially by the results of our ERP study, where we compared electrophysiological responses to simple, neutral sounds, neutral faces, and faces and voices presented together (see Figure 8) (Liu et al., "Emotional"). First, there is a clear distinction between voice only stimuli, with a clear ERP morphology consisting of N100 and P200 components, typically associated with voice recognition and evident at central and frontal scalp locations. The N100 and P200 to voice only stimuli are absent from the occipital and parietal locations suggesting that brain regions involved in their generation are situated at temporal locations and are not sensitive to visual input. In contrast, there is a highly similar morphology of face only and voice-face stimuli, with most robust ERP components registered over central and frontal scalp locations, but also extending to occipital and parietal sites. Even though ERP scalp distribution in no way reflects one to one correspondence between brain sources and scalp-registered potentials, the broad distribution of ERPs to voice-face stimuli suggests that multiple brain sources contribute to our ability to receive integrated messages from both face and voice. At the same time, the significant difference between face only and voice-face stimuli observed for the N250 ERP component is in agreement with the temporal contributions, such as from STS and pSTG, to voice-face processing.

This very brief overview of how signals from speech and face are integrated into a coherent message highlight at least two observations: the speed with which critical information is extracted from the incoming physical signal and the complexity of brain regions implicated in the efficient and meaningful interpretation of that information.

At the same time, this review highlights how inefficient or abnormal processes within brain networks involved in speech and face processing may contribute to profound deficits and difficulties in individuals impacted by them. While several clinical conditions are believed to be related to neurocognitive deficits in the brain systems discussed above, I will discuss schizophrenia as an example of severe impairments rooted in the abnormal processing of speech and face information.

7. Deficits in prosody processing in schizophrenia and its consequences

Schizophrenia is a severe disease with genetic and neurodevelopmental components characterized by thought disorder expressed in abnormal language use and distortions in reality perception. These distortions include complex delusions, unfounded fears, as well as visual and auditory hallucinations which result in 'perceiving' images and hearing voices in the absence of a visual or an auditory input. It is believed that many of these symptoms arise from abnormalities in brain regions that are involved in both speech and face processing.

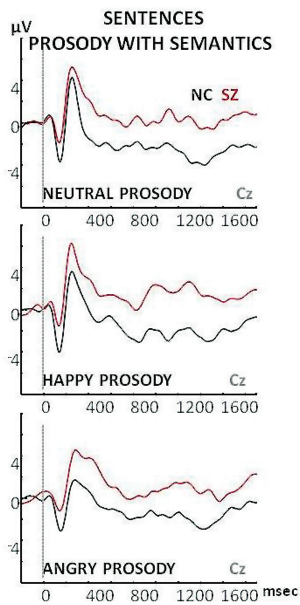

Figure 9: Grand average waveforms to sentence prosody in 18 schizophrenia (SZ) and 18 healthy control (NC) subjects to neutral, happy and angry prosody (semantic information present)

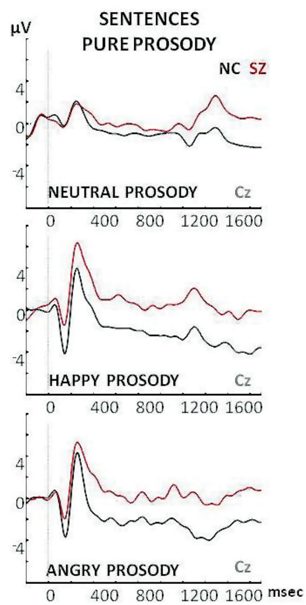

Figure 10: Grand average waveforms to sentence prosody in 18 schizophrenia (SZ) and 18 healthy control subjects (NC) to neutral, happy and angry prosody (with semantic information removed)

Models of prosody processing abnormality in schizophrenia: It has been demonstrated that individuals with a schizophrenia diagnosis experience difficulties in prosody processing at all stages of analysis. We have demonstrated that both the initial stages of sensory analysis as indexed by the N100 ERP component, and categorization of a sensory input into higher order entities, as indexed by the P200 component, are abnormal in chronic schizophrenia (Pinheiro et al., "Interactions"; Pinheiro et al., "Abnormal").

As can be seen in Figures 3, 4, 9 and 10, these deficits are especially prominent for emotional prosody; here, happy and angry voices. Furthermore, these deficits exist both in response to sentences with and without semantic content underscoring a unique nature of difficulties in processing prosodic information in schizophrenia.

The deficits documented with the ERP data and manifesting at the earliest stages of speech processing are further corroborated by fMRI studies that identify processing deficits in both temporal and frontal brain regions involved in prosody analysis, including in the orbitofrontal cortex associated with linking sensory data to meaning. One of the models of prosody dysfunction' in schizophrenia suggests that the prosodic abnormality is related to impairment at both early, sensory stages and at later, higher order cognition stages and underscores an interactive relationship between these two (Leitman et al.). Furthermore, these initial deficits are exacerbated by deficits in attention and working memory.

The consequences of these impairments are serious. They contribute to emotional withdrawal and affective blunting, as well as developing suspiciousness and paranoia. For example, individuals with schizophrenia tend to misinterpret speakers' intentions by assigning a hostile intention to happy or neutral prosody and exaggerating a hostile intent for angry prosody (e.g., Lin, Ding and Zhang). Furthermore, as mentioned above, there is a close relationship between emotion recognition and theory of mind capacities, with the latter impacted by schizophrenia irrespective of prosodic deficits. Thus, while performance on a prosody task may seem of a distant relevance to the ability to function in a real world, it can be argued that, as described above, it has profound consequences for successful social functioning both in professional and private spheres of life. Prosodic deficits can lead to constructing a threatening model of the world, with people intent to harm an individual. Left untreated, it can lead to withdrawal from society and significant suffering.

8. Auditory hallucinations as a special case of auditory and social cognition deficits

Auditory hallucinations are severely disturbing auditory phenomena that can be deeply unsettling for those who experience them. They are not unique to schizophrenia and are reported in a number of clinical conditions. Mostly, but not always, they are critical or threatening towards a person who experiences them. They vary in their manifestations. They can be experienced as a single voice, or several voices conversing, and the voices' gender may or may not align with the gender of their hearer. The critical hallmark of voice hearing is that it is experienced in the absence of an outside auditory source. Thus, the question becomes not of what kind of distortions occurred to the auditory source, but rather what kind of neurocognitive events, in what kind of brain regions, contributed to this phenomenon which, to those who experience it, is often as real as a conversation with another person or persons.

Early models suggested that auditory hallucinations are a result of source misattribution during self-generated thought and inner speech, where patients attributed the source to others instead of self (Frith et al.; Frith and Done). In these models, it was assumed that people think using silent language and when this silent language produced in the temporal and frontal cortices is misattributed, it is experienced as external speech. More recent models have suggested a breakdown in the network of brain regions involved in auditory hallucinations, whereby the over-activated auditory perceptual regions in the temporal cortex are coupled with a lack of inhibitory control exhibited by the executive function regions in the frontal cortex (Hugdahl). In addition, there have been proposals that link auditory hallucinations to abnormalities in the motor speech and language regions (Allen et al.; Jardri et al.). Here, the assumption is that perceptual regions of the temporal cortex are abnormally active, and since the frontal regions do not suppress this activation efficiently, the outputs of the temporal regions are being imbued with meaning.

Finally, there are models which conceptualize auditory hallucinations as a disturbance in agency, where agency refers to a sense of self as contrasted with other agents or people (e.g., Brent et al.; Holt et al.). The STG and medial prefrontal cortex (MPFC) have been shown to be a part of self-referential network in both healthy (Jenkins and Mitchell; Kelley et al.) and individuals with schizophrenia (Brent et al.; Larivière et al.).

It is quite likely that a complete model of auditory hallucinations should be a syncretic one; there is evidence that indeed actively hallucinating patients relative to non-hallucinating patients show increased activation in the STG, in the middle temporal gyrus (MTG), and in the parietal regions, such as the temporo-parietal junction (TPJ), angular gyrus, and inferior parietal lobule (Jardri et al.; Thoma et al.). In addition, resting state connectivity between MPFC and posterior cingulate cortex (PCC), two hubs of the default mode network involved in self-reflection, is abnormally high in patients experiencing auditory hallucinations.

As a phenomenon, auditory hallucinations illustrate how effective communication depends on the fidelity of analyses of external auditory signals but also how brain regions dedicated to this task in the course of evolution play a crucial role in connecting us with an outside world. This connection depends critically on their unincumbered function. In the case of auditory hallucinations, abnormal function of brain regions dedicated to speech processing can produce hearing experiences that mimic speech and yet do not reflect outside input. Most recent efforts that use fMRI based neurofeedback techniques suggest that direct impacting brain regions involved in auditory processing, and in self-other distinctions, can in fact help normalize the function of these regions and lead to reductions in auditory hallucinations (Okano et al.; Bauer et al.).

9. Deficits in face processing and in face and voice processing in schizophrenia

The difficulties in the processing of the auditory signal, in terms of the veridical analysis of the sensory data, and in terms of generating faulty data due to abnormalities in the brain structures dedicated to such analyses are not limited to speech. There is ample evidence that both analyses of the sensory data derived from faces, and the higher-order operations based on these data, are also abnormal in schizophrenia (Rubin et al.; Onitsuka et al.; Salisbury et al.; Feuerriegel et al.). This deficit extends to the simultaneous processing of faces and voices (Liu et al., "Simultaneous"). As discussed above, the study conducted by our group showed that both face only and voice-face stimuli were impacted by schizophrenia (see Figure 11), even though this impact differed as a function of brain regions involved.

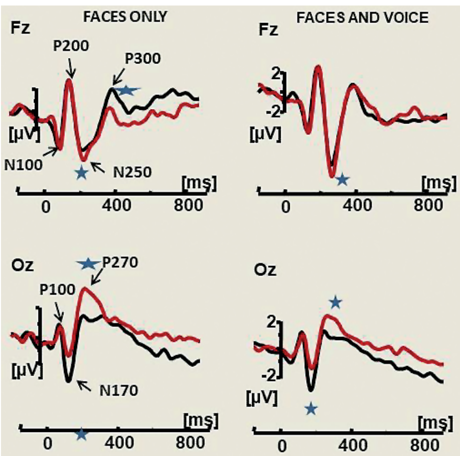

Figure 11: Abnormal ERP responses to neutral faces, and to face-voice stimuli in patients with schizophrenia (red line) relative to healthy control individuals (black line): both sensory processes (P100 and N170) associated with structural face processing, as well as P270 associated with categorization processes, were found affected

10. Conclusions

Our world is accessible to us through our senses, which have uniquely evolved in humans to allow to engage with the world in a purposeful way. Arguably, information conveyed by our senses is species-specific and the world is experienced differently by humans, fish, octopus, and dogs (even though there is some evidence that domesticated animals like dogs show some remarkable abilities to share human perception of the outside world; for example, in understanding some words—e.g., Andics and Miklosi; Pepperberg). I have focused, in this brief review, on physical signals conveyed by speech and faces, as it is recognized that these two mediums provide the richest sources of information. Using speech, we describe the world according to our best understanding, and long before we could write, we told stories about where we came from and where we were going. The way we said it made often all the difference. With prosody, we signaled love, hate, admiration and contempt, and used noises made by our vocal system to establish our place in society. It has been recognized that the use of speech and prosody, as well as their appropriate decoding are a part of what we call social cognition and is vital to communication in any given society, and indeed indispensable to its creation. Our facial expressions provide another layer of complexity and an additional source of information about a social world around us. Our nearly instantaneous access to the wealth of data derived from purely physical properties of auditory and visual input belies the fact that the brain processes involved in delivering what we receive as a message are immensely complex. They rely on a network of regions which provide initial analyses of sensory data, a scaffolding for the interface between the sensory and the abstract, and assigning a final meaning, often modulated by additional inputs from brain regions involved in attention, memory, self-reflection, and self–other distinctions. As briefly discussed in the context of schizophrenia, the seemingly simple deficits experienced by non-neurotypical individuals have profound consequences for their ability to function in a social world. As a species, we are a product of our evolutionary history, inhabiting the world between the physical and non-physical, both creators and created, defined by biophysical properties of our brains but also with aspirations, and some capabilities, to change them.

References

Allen, P. et al. 2008. "The Hallucinating Brain: A Review of Structural and Functional Neuroimaging Studies of Hallucinations." *Neuroscience & Biobehavioral Reviews* 32.1. 175–91.

Allison, T. et al. 1999. "Electrophysiological Studies of Human Face Perception. I: Potentials Generated in Occipitotemporal Cortex by Face and Non-face Stimuli." *Cerebral Cortex* 9. 415–30.

Andics, A. and Á. Miklósi. 2018. "Neural Processes of Vocal Social Perception: Dog-human Comparative fMRI Studies." *Neuroscience & Biobehavioral Reviews* 85. 54–64.

Basnáková, J. et al. 2014. "Beyond the Language Given: The Neural Correlates of Inferring Speaker Meaning." *Cerebral Cortex* 24. 2572–78.

Bauer C. C. C. et al. 2020. "Real-time fMRI Neurofeedback Reduces Auditory Hallucinations and Modulates Resting State Connectivity of Involved Brain Regions. Part 2: Default Mode Network—Preliminary Evidence." *Psychiatry Research* 284. Art. 112770.

Bhaya-Grossman, I. and E. F. Chang. 2022. "Speech Computations of the Human Superior Temporal Gyrus." *Annual Review of Psychology* 4.73. 79–102.

Brent, B. K. et al. 2014. "Neural Responses During Social Reflection in Relatives of Schizophrenia Patients: Relationship to Subclinical Delusions." *Schizophrenia Research* 157.1–3. 292–98.

Chai, L. R. et al. 2016. “Functional Network Dynamics of the Language System.” *Cerebral Cortex* 26.11. 4148–59.

Chandrasekaran, C and A. A. J. Ghazanfar. 2009. “Different Neural Frequency Bands Integrate Faces and Voices Differently in the Superior Temporal Sulcus.” *Neurophysiology* 101.2. 773–88.

Crowley, K. E. and I. M. Colrain. 2004. “A Review of the Evidence for P2 Being an Independent Component Process: Age, Sleep and Modality.” *Clinical Neurophysiology* 115.4. 732–44.

Cruz, S. et al. 2022. “Neural and Psychophysiological Correlates of Social Communication Development: Evidence from Sensory Processing, Motor, Cognitive, Language and Emotional Behavioral Milestones across Infancy.” *Applied Neuropsychology: Child* 11.2. 158–77.

Fedorenko, E. and S. L. Thompson-Schill. 2014. “Reworking the Language Network.” *Trends in Cognitive Sciences*18.3. 120–26.

Feuerriegel, D. et al. 2015. “The N170 and Face Perception in Psychiatric and Neurological Disorders: A Systematic Review.” *Neurophysiology* 126.6. 1141–58.

Friederici, A. D. 2017. “Evolution of the Neural Language Network.” *Psychonomic Bulletin & Review* 24.1. 41–47.

Friederici, A. D. and S. M. E. Gierhan. 2013. “The Language Network.” *Current Opinion Neurobiology* 23.2. 250–54.

Friederici, A. D. et al. 2003. “The Role of Left Inferior Frontal and Superior Temporal Cortex in Sentence Comprehension: Localizing Syntactic and Semantic Processes.” *Cerebral Cortex* 13.2. 170–77.

Frith, C. D. and D. J. Done. 1988. “Towards a Neuropsychology of Schizophrenia.” *British Journal of Psychiatry* 153. 437–43.

Frith, C. D. et al. 1992. “PET Imaging and Cognition in Schizophrenia.” *Journal of the Royal Society of Medicine* 85.4. 222–24.

Garcia-Larrea, L., A. C. Lukaszewicz and F. Mauguiere. 1992. „Revisiting the Oddball Paradigm: Non-target Vs Neutral Stimuli and the Evaluation of ERP Attentional Effects.” *Neuropsychologia* 30.8. 723–41.

Getz, L. M. and J. C. Toscano. 2021. “The Time-course of Speech Perception Revealed by Temporally-sensitive Neural Measures.” *Wiley Interdisciplinary Reviews: Cognitive Science* 12.2. Art. 1541.

Hagoort, P. 2019. “The Neurobiology of Language Beyond Single-word Processing.” *Science* 366.6461. 55–58,

Holt, D. J. et al. 2011. “An Anterior-to-posterior Shift in Midline Cortical Activity in Schizophrenia During Self-reflection.” *Biological Psychiatry* 69.5. 415–12.

Hugdahl, K. 2015. “Auditory Hallucinations: A Review of the ERC ‘VOICE’ Project.” *World Journal of Psychiatry* 5.2. 193–209.

Iidaka, T. et al. 2006. “Hemodynamic and Electrophysiological Relationship Involved in Human Face Processing: Evidence from a Combined fMRI-ERP Study.” *Brain Cognition* 60.2. 176–86.

Jardri, R. et al. 2011. “Cortical Activations During Auditory Verbal Hallucinations in Schizophrenia: A Coordinate-based Meta-analysis.” *American Journal of Psychiatry* 168.1. 73–81.

Jenkins, A. C. and J. P. Mitchell. 2011. “Medial Prefrontal Cortex Subserves Diverse Forms of Self-reflection.” *Social Neuroscience* 6.3. 211–18.

Kanske, P., J. Plitschka and S. A. Kotz. 2011. „Attentional Orienting Towards Emotion: P2 and N400 ERP Effects.” *Neuropsychologia* 49.11. 3121–29.

Kelley, W. M. et al. 2002. “Finding the Self? An Event-related fMRI Study.” *Journal of Cognitive Neuroscience* 14.5. 785–94.

Kotz, S. A. and S. Paulmann. 2007. “When Emotional Prosody and Semantics Dance Cheek to Cheek: ERP Evidence.” *Brain Research* 1151. 107–18.

Kreifelts, B et al. 2007. “Audiovisual Integration of Emotional Signals in Voice and Face: An Event-related fMRI Study.” *Neuroimage* 37.4. 1445–56.

Kutas, M. and K. D. Federmeier. 2000. “Electrophysiology Reveals Semantic Memory Use in Language Comprehension.” *Trends in Cognitive Sciences* 4.12. 463–70.

—. 2011. “Finding Meaning in the N400 Component of the Event-related Brain Potential (ERP).” *Annual Review of Psychology* 62. 621–47.

Lang, P. J., M. M. Bradley and B.N. Cuthbert. 2008. *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-8*. Gainesville, FL: University of Florida.

Larivière, S. et al. 2017. “Altered Functional Connectivity in Brain Networks Underlying Self-referential Processing in Delusions of Reference in Schizophrenia.” *Psychiatry Research* 263. 32–43.

Leitman D. I. et al. 2010. “Sensory Deficits and Distributed Hierarchical Dysfunction in Schizophrenia.” *American Journal of Psychiatry* 167.7. 818–27.

Leitman, D. I. et al. 2011. “Preliminary Evidence of Pre-attentive Distinctions of Frequency-modulated Tones that Convey Affect.” *Frontiers in Human Neuroscience* 5. Art. 96.

Lenz, D. et al. 2007. “What’s That Sound? Matches with Auditory Long-term Memory Induce Gamma Activity in Human EEG.” *International Journal of Psychophysiology* 64.1. 31–38.

Lin, Y., H. Ding and Y. Zhang. 2018. “Emotional Prosody Processing in Schizophrenic Patients: A Selective Review and Meta-analysis.” *Journal of Clinical Medicine* 7.10. Art. 363.

Liu, L. et al. 2022. “How Tone, Intonation and Emotion Shape the Development of Infants’ Fundamental Frequency Perception.” *Frontiers in Psychology* 13. Art. 906848.

Liu, T. et al. 2012. “Electrophysiological Insights into Processing Nonverbal Emotional Vocalizations.” *Neuroreport* 23.2. 108–12.

Liu, T. et al. 2012. “Emotional Cues During Simultaneous Face and Voice Processing: Electrophysiological Insights.” *PLOS One* 7.2. Art. e31001.

Liu, T. et al. 2016. “Simultaneous Face and Voice Processing in Schizophrenia.” *Behavioral Brain Research* 305. 76–86.

Mar, R. A. 2011. “The Neural Bases of Social Cognition and Story Comprehension.” *Annual Review of Psychology* 62. 103–34.

Mars, R. B. et al. 2012. “Connectivity-based Subdivisions of the Human Right Temporoparietal Junction Area: Evidence for Different Areas Participating in Different Cortical Networks.” *Cerebral Cortex* 22.8. 1894–903.

Mitchell, R. L. and L. H. Phillips. 2015. “The Overlapping Relationship Between Emotion Perception and Theory of Mind.” *Neuropsychologia* 70. 1–10.

Mitchell, R. L. and S. L. Rossell. 2014. “Perception of Emotion-related Conflict in Human Communications: What Are the Effects of Schizophrenia?” *Psychiatry Research* 220.1–2. 135–44.

Niedenthal, P. M. et al. 2010. “The Simulation of Smiles (SIMS) Model: Embodied Simulation and the Meaning of Facial Expression.” *Behavioral Brain Science* 33.6. 417–33.

Okada, K. et al. 2010. “Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech.” *Cerebral Cortex* 20.10. 2486–95.

Okano, K. et al. 2020. “Real-time fMRI Feedback Impacts Brain Activation, Results in Auditory Hallucinations Reduction. Part 1: Superior Temporal Gyrus. Preliminary Evidence.” *Psychiatry Research* 286. Art. 112862.

Onitsuka, T. et al. 2020. “Altered P3a Modulations to Emotional Faces in Male Patients with Chronic Schizophrenia.” *Clinical EEG and Neuroscience* 51.4. 215–21.

Paulmann, S. and S. A. Kotz. 2008. “An ERP Investigation on the Temporal Dynamics of Emotional Prosody and Emotional Semantics in Pseudo- and Lexical-sentence Context.” *Brain and Language* 105.1. 59–69.

Paulmann, S., S. Seifert and S. A. Kotz. 2009. "Orbito-frontal Lesions Cause Impairment During Late but not Early Emotional Prosodic Processing." *Social Neuroscience* 5.1. 59–75.

Pepperberg, I. M. 2021. "Nonhuman and Nonhuman-Human Communication: Some Issues and Questions." *Frontiers in Psychology* 12. Art. 64784.

Perrodin, C. et al. 2015. "Natural Asynchronies in Audiovisual Communication Signals Regulate Neuronal Multisensory Interactions in Voice-sensitive Cortex." *Proceedings of the National Academy of Sciences* 112.1. 273–78.

Pinheiro, A. P. et al. 2013. "Interactions Between Mood and the Structure of Semantic Memory: Event-related Potentials Evidence." *Social Cognitive and Affective Neuroscience* 8.5. 579–94.

Pinheiro, A. P. et al. 2015. "Abnormal Interactions Between Context, Memory Structure, and Mood in Schizophrenia: An ERP Investigation." *Psychophysiology* 52.1. 20–31.

Powell, L. J., H. L. Kosakowski and R. Saxe. 2018. "Social Origins of Cortical Face Areas." *Trends in Cognitive Sciences* 22.9. 752–63.

Rizzolatti, G. and C. Sinigaglia. 2010. "The Functional Role of the Parieto-frontal Mirror Circuit: Interpretations and Misinterpretations." *Nature Reviews Neuroscience* 11. 264–74.

Rubin, L. H. et al. 2022. "Real-time Facial Emotion Recognition Deficits Across the Psychosis Spectrum: A B-SNIP Study." *Schizophrenia Research* 243. 489–99.

Sadeh, B. and G. Yovel. 2010. "Why Is the N170 Enhanced for Inverted Faces? An ERP Competition Experiment." *Neuroimage* 53.2. 782–89.

Salisbury, D. F. et al. 2019. "Neutral Face and Complex Object Neurophysiological Processing Deficits in Long-term Schizophrenia and in First Hospitalized Schizophrenia-spectrum Individuals." *International Journal of Psychophysiology* 145. 57–64.

Schirmer, A. and S. A. Kotz. 2006. "Beyond the Right Hemisphere: Brain Mechanisms Mediating Vocal Emotional Processing." *Trends in Cognitive Sciences* 10.1. 24–30.

Spunt, R. P. and R. Adolphs. 2017. "A New Look at Domain Specificity: Insights from Social Neuroscience." *Nature Reviews Neuroscience* 18.9. 559–567.

—. 2019. "The Neuroscience of Understanding the Emotions of Others." *Neuroscience Letters* 693. 44–48.

Thoma, R. J.et al. 2016. "Functional MRI Evaluation of Multiple Neural Networks Underlying Auditory Verbal Hallucinations in Schizophrenia Spectrum Disorders." *Frontiers in Psychiatry* 29.7. Art. 39.

Yovel, G. 2016. "Neural and Cognitive Face-selective Markers: An Integrative Review." *Neuropsychologia* 83. 5–13.